



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Complete nucleotide sequence of the murine H-2k k gene. Comparison of three H-2K locus alleles

Citation for published version:

Arnold, B, Burgert, H, Archibald, AL & Kvist, S 1984, 'Complete nucleotide sequence of the murine H-2k k gene. Comparison of three H-2K locus alleles', *Nucleic Acids Research*, vol. 12, no. 24, pp. 9473-9487.
<https://doi.org/10.1093/nar/12.24.9473>

Digital Object Identifier (DOI):

[10.1093/nar/12.24.9473](https://doi.org/10.1093/nar/12.24.9473)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Nucleic Acids Research

Publisher Rights Statement:

I
RL
Press
Limited,
Oxford,
England.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Complete nucleotide sequence of the murine H-2K^k gene. Comparison of three H-2K locus alleles

Bernd Arnold⁺, Hans-Gerhard Burgert[§], Alan L. Archibald[§] and Sune Kvist^{§*}

European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG

Received 2 November 1984; Revised and Accepted 30 November 1984

ABSTRACT

We have determined the DNA sequence of the H-2K^k gene of the mouse major histocompatibility complex (MHC). Comparison on the nucleotide and protein level of three H-2K alleles (K^k, K^b and K^d) reveals a high degree of homology, in particular between the K^b and K^k alleles. Differences between the two latter antigens are almost exclusively confined to the $\alpha 1$ and $\alpha 2$ domains. At nine positions in the extracellular part of the molecules we have found allele-specific amino acids. Interestingly, 78% of these residues are either polar or carry hydroxyl-groups. This makes it likely that they are exposed on the surface of the molecules and might then be part of antigenic determinants. We have also identified potentially allele-specific nucleotide sequences of the K genes which might be used as specific DNA probes.

INTRODUCTION

The murine major histocompatibility complex (MHC) encodes the transplantation antigens H-2K, D and L (1,2). These class I antigens consist of a heavy chain (MW 45000) non-covalently associated with β_2 -microglobulin (MW 12000). The heavy chain is an integral membrane protein with the largest part of the molecule exposed on the cell surface and the remaining one fourth of the protein spanning the plasma membrane and protruding on the cytoplasmic side. The extracellular part of the molecule can be divided into three domains ($\alpha 1$, $\alpha 2$ and $\alpha 3$) each comprising approximately 90 amino acids. The primary structure of the protein has been determined for the H-2K^b molecule (3). One of the most striking features of the class I antigens is their high degree of polymorphism. More than 50 alleles at each of the K and D loci have so far been identified (2).

The class I histocompatibility antigens play an important role in the T cell mediated immune response. During viral infection cytotoxic T lymphocytes show a specificity not directed against the viral proteins themselves but against the viral antigens in the context of self-determinants contained on H-2 class I antigens. The role of the transplantation antigens seems to be to restrict the recognition so that only the infected cells are lysed. This phenomenon has been called H-2 restriction of cytolytic T lymphocytes (4,5,6).

During the past few years the understanding of the molecular biology of the H-2 complex has advanced dramatically. By using recombinant DNA technology, several laboratories have isolated and characterized class I genes (7,8,9,10,11) and a molecular map of the H-2 complex has been constructed (12).

The H-2 class I genes so far analysed display a high degree of homology. They all have 8 exons which correlate with the domains of the corresponding antigen. The first exon encodes the signal sequence, known to be of importance for translocation of the newly synthesized polypeptide through the membrane of the rough endoplasmic reticulum (13). Exons 2,3 and 4 encode the three extracellular domains of the antigen, $\alpha 1$, $\alpha 2$ and $\alpha 3$, respectively. Exon 5 encodes the transmembrane domain, whereas exons 6, 7 and 8 code for the cytoplasmic domain and the 3' untranslated region. The overall length of an H-2 class I gene is approximately 5000 nucleotides (for a review see 14).

Two alleles of the H-2K locus, K^b and K^d , have already been characterized (11,10). In this report we describe the isolation and characterization of a third allele of the H-2K locus, the K^k gene. This allows us to make a comparison, on the protein as well as on the DNA level, of three alleles from an H-2 class I locus.

MATERIALS AND METHODS

Isolation of lambda phages containing H-2 sequences and Southern blot hybridization

The H-2^k lambda phage library was a gift from H. Lehrach and A.-M. Frischauf. The library was constructed from DNA of the

strain B10.BR partially digested with the restriction enzyme Sau 3A1 and cloned into the Bam HI site of the EMBL3 vector (15). Phage clones containing genomic DNA were identified by plaque hybridization (16). Phage DNA was prepared as described (17). Southern blot analyses of phage DNA were carried out as described elsewhere (10).

DNA restriction mapping and sequence analysis

Restriction maps of lambda phage clones were constructed either as previously described (18) or by using two or several restriction enzymes in consecutive reactions. For DNA sequence analysis we have used the subcloning deletion method (19). Overlapping clones were sequenced by the procedure of Maxam and Gilbert (20).

Cell culture and DNA-mediated gene transfer

The introduction of the H-2K genes into the fibroblast cell line 1T 22-6 was done as previously described (21,23). Selection of clones expressing H-2K antigens was done using the neophosphotransferase gene system (22). These experiments have been published in detail elsewhere (23).

Other methods

Immunoprecipitations and sodium dodecyl sulphate polyacrylamide gel electrophoretic analyses were performed essentially as described before (24). For the detection of the H-2K^k antigen the monoclonal antibody H100-27R55 was used (25).

Materials

Restriction enzymes and DNA modifying enzymes were purchased from Boehringer, Mannheim, FRG. (α -³²P)dNTPs (10mCi/ml, 3000Ci/mmol), (γ -³²P)ATP (10mCi/ml, 5000Ci/mmol) and ³⁵S-methionine were from Amersham, U.K. GeneScreen filters were from New England Nuclear. Protein A-sepharose was from Pharmacia Fine Chemicals (Uppsala, Sweden). Geneticin (G-418) was purchased from GIBCO.

RESULTS

Isolation of genomic lambda clones containing H-2 specific sequences

A genomic library, constructed from liver DNA of the mouse strain B10.BR (H-2^k) and the vector EMBL 3 (15), was screened with the cDNA probe pH-2^d-4

(26). This probe has its 5' end at a position corresponding to the 3' end of exon 3 of the K^d antigen and continues through the $\alpha 3$ domain (exon 4), the membrane and the cytoplasmic domains and ends after a stretch of approximately 50 adenosine residues at the 3' end. Since it contains the highly conserved region of the $\alpha 3$ domain, we expected it to recognize most, if not all, H-2 class I genes in the mouse genome. Approximately, 5×10^5 individual lambda clones were screened, of which 112 were found to hybridize specifically to the H-2 probe. The number of clones screened corresponds to roughly 3 times the haploid mouse genome. The 35-40 genes detected per haploid genome are in agreement with data reported before (12).

From 40 of the 112 clones picked, DNAs were prepared and analysed in pools (5-10 individual clones per pool) for their ability to hybridize to an H-2K locus specific probe (pH-2^d-5b). This probe originates from the 3' untranslated region of the K^d gene as has been characterized previously (27).

DNAs from the different pools were subjected to digestion with the restriction enzyme Bgl II, the fragments were separated on an agarose gel, transferred to a GeneScreen filter and hybridized with the radioactively labelled probe pH 2^d-5b (27). Two of the six pools examined contained a single hybridizing fragment. The sizes of these fragments were 3.2 Kb and 2.4 Kb, respectively. In a Southern blot experiment with total genomic B10.BR DNA we were unable to detect these two bands, most likely for the following reasons: First, each band seems to represent a single copy gene and secondly, the probe pH-2^d-5b contains approximately 75% A+T residues which makes hybridization experiments with total DNA difficult. DNA from the individual clones of pools 4 and 2 were then analysed in a similar way. Two clones, termed 6D and 7A, were shown to contain DNA sequences complementary to the probe pH-2^d-5b. These two clones were considered to be candidates for the authentic H-2K^k gene and were further analyzed by two independent techniques, described below.

Restriction mapping and DNA-mediated gene transfer of H-2K genes

The restriction maps for clones 6D and 7A reveal a high degree of homology but differ at three positions in the 3' flanking region (Figure 1). This indicates that the clones contain two individual H-2 genes. By using 3' and 5' H-2 specific DNA probes in Southern blot experiments, we could show that clone 6D does not contain a complete H-2 gene. This clone does not hybridize to the 5' specific probe containing a sequence encoding the $\alpha 1$

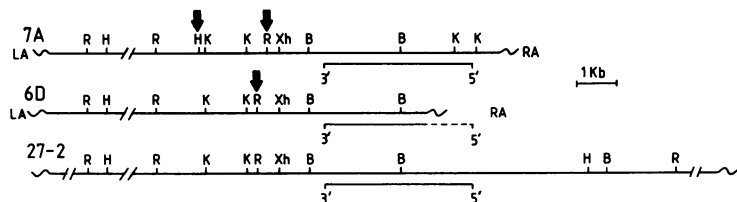


Figure 1. Restriction maps of the lambda clones 7A, 6D and the cosmid clone 27-2. The vertical arrows denote the three differences between clones 7A and 6D. The horizontal bars underneath each of the restriction maps show the extent of the H-2 gene. For clone 6D the 5' end of the gene is missing and this is indicated by the dashed bar. LA and RA mean the left and the right arms, respectively, of the lambda phage DNA. The sign ~ denotes lambda phage or cosmid DNA sequences. Enzymes used were: B, BamHI; H, Hind III; K, Kpn I; R, EcoRI; Xh, Xho I.

domain of the H-2 K^d gene. Also the homology between the two clones indicates that the 5' end of the gene in clone 6D is missing. This is the result of the cloning procedure used. The H-2 gene in clone 6D contains all the sequences encoding exon 4 to exon 8 as well as the 3' non-coding region but lacks the first three exons upstream of the Bam HI site. Clone 7A contains a complete H-2 gene according to restriction mapping and Southern blot analysis.

In order to find a complete gene corresponding to the gene in clone 6D, we examined the other clones originally selected to contain H-2 sequences. None of these clones contained a complete gene corresponding to clone 6D or a gene which with certainty could provide the authentic 5' end of the gene in clone 6D. We then examined a cosmid clone, 27-2 (28) (generously provided by M. Steinmetz) which originates from a cosmid library of DNA from the mouse strain AKR ($H-2^k$). Clone 27-2 was selected on the basis of its hybridization to the K-locus specific probe ($pH-2^d-5b$) and therefore could be considered a candidate for containing the H-2 K^k gene. By comparing the restriction map of cosmid clone 27-2 with clone 6D and 7A, it is clear that clones 27-2 and 6D are identical at their 3' ends and that the former, in addition, contains the sequences upstream of the Bam HI site of clone 6D (Figure 1).

By using DNA-mediated gene transfer experiments with each of the clones we could demonstrate that only DNA of clone 27-2 gave rise to cells expressing the authentic H-2K^k gene (23). Transfected cells were analysed with both sodium dodecyl sulphate polyacrylamide gel electrophoresis and fluorescence activated cell sorter analysis.

From the experiments described above, we conclude the following: (i) Clone 27-2 contains the H-2K^k gene of the AKR mouse. (ii) Clone 6D probably contains the 3' end of the H-2K^k gene of the mouse B10.BR. (iii) Clone 7A contains the DNA sequence of a complete H-2 gene bearing at its 3' end the H-2K locus specific sequence. This clone contains an H-2 gene from the B10.BR genome but it is not the authentic H-2K^k gene.

Amino acid and nucleotide comparison of three H-2K class I genes

The 10.5 Kb Eco RI fragment of cosmid 27-2 was recloned into the unique Sal I site of pBR 322 by using Sal I linkers. Plasmid clones containing the 10.5 Kb fragment were selected and for one of these, clone 27-2-86, a partial restriction map was constructed (Figure 2). In order to determine the nucleotide sequence of the K^k gene, clone 27-2-86 was subjected to the subcloning deletion method (19). We have used the Cla I linker to introduce the corresponding restriction site at different positions along the DNA as described elsewhere (23). The DNA sequence was determined and aligned with those of the K^b and K^d genes (11, 10) to identify the exon-intron boundaries (Figures 2 and 3).

The nucleotide sequence of the K^k gene shows extensive homology with other H-2 class I genes (Figure 3). It is encoded by 8 exons of varying lengths separated by introns. Two transcription promoting elements are found upstream of the gene. The TATA box is located 55 nucleotides upstream of the initiation codon ATG and the CCAAT-box is found another 27 nucleotides upstream. Both these sequences have been postulated to play an important role in gene regulation (29). Both sequences are located at almost identical position in the K^b and K^d genes.

We have compared the three H-2K alleles with respect to nucleotide homology in different regions of the genes (Table I). Exons

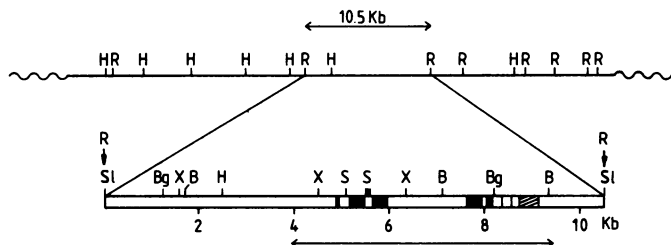


Figure 2. Restriction map of the 10.5 Kb fragment containing the H-2K^k gene of cosmid 27-2.

The 10.5 Kb EcoRI fragment was cloned into the Sal I site of pBR322 by using Sal I linkers. Filled boxes denote exons and open boxes denote introns and flanking sequences. The dashed box shows the 3' untranslated region. The horizontal arrow underneath the detailed restriction map indicates the region for which the DNA sequence has been determined. Enzymes used were as in Figure 1 and in addition: Bg, Bgl II; S, Sma I; Sl, Sal I; X, Xba I.

two and three (encoding the $\alpha 1$ and $\alpha 2$ domains) display the lowest degree of homology (88-92%) whereas the cytoplasmic exons show the most extensive homology (97-100%). The K^b and K^k genes are slightly more homologous to each other than to the K^d gene. There are only nine nucleotide differences between K^b and K^k in exons 4 to 8 and they differ by only 13 nucleotides from the start of the $\alpha 3$ domain (exon four) to the termination codon. Thus, K^b and K^k sequences are 99% homologous (13 differences out of a total of 1089 nucleotides) in this region. This degree of conservation in coding as well as in non-coding sequences is remarkable.

The overall homology on the protein level between the three antigens is for K^b-K^k 90%, for K^d-K^k 86% and for K^b-K^d 83%. Thus, the K^b and K^k antigens are much more similar to each other than to the K^d antigen. Of the total of 35 amino acid differences between K^b and K^k, 27 are contained in the $\alpha 1$ and $\alpha 2$ domains (Figure 4). Only two differences are found in the $\alpha 3$ domain (Figure 4, positions 191 and 225). This is a surprisingly low number, considering that K^d and K^k have 11 amino acid substitutions and K^b and K^d have 12 substitutions in the same region. Amino acids unique at certain positions for each allele are found

[illegible]

TABLE I: Nucleotide sequence homology* of exons and introns of three H-2 K alleles

Alleles compared	<u>Exon:</u>	<u>I</u>	<u>II</u>	<u>III</u>	<u>IV</u>	<u>V</u>	<u>VI-VIII</u>	
K ^k -K ^b		94	90	91	98	96	100	
K ^k -K ^d		95	88	92	94	96	97	
K ^d -K ^b		92	90	89	94	91	97	
	<u>Introns:</u>	<u>I</u>	<u>II</u>	<u>III</u>	<u>IV</u>	<u>V</u>	<u>VI</u>	<u>VII</u>
K ^k -K ^b		95	95	- ^s	100	100	100	96
K ^k -K ^d		95	96	94	95	96	98	98
K ^d -K ^b		93	96	- ^s	95	96	98	98

* The homology is given as percentage

§ Sequence information for the third intron of the H-2K^b gene is not complete

in nine positions in the three extracellular domains (Figure 4). Five of these are contained in the $\alpha 1$ domain and of these all except one are present in the hypervariable cluster at positions 62-83. Another three are scattered in the $\alpha 2$ domain at positions 99, 156 and 173. The $\alpha 3$ domain contains the last allele-specific residue at position 191.

The K^k and K^b antigens each have two glycosylation sites, one in the $\alpha 1$ domain at position 86 and the second in the $\alpha 2$ domain at position 176 (Figure 4). The K^d antigen has an additional carbohydrate group in the $\alpha 3$ domain at residue 256. At this position the K^b and K^k polypeptides contain tyrosine residues.

Figure 3. Complete DNA sequence of the H-2K^k gene.

The sequence was determined as previously described (19) in combination with the Maxam and Gilbert procedure (20). The deduced amino acid sequence is shown on top of the DNA sequence. Transcriptional promoting elements are underlined. The first nine amino acids for exon 8 are given in italics because of the uncertainty over which acceptor site is used.

α1-DOMAIN										10	20													
H-2	K ^k	Gly	Pro	His	Ser	Leu	Arg	Tyr	Phe	His	Thr	Ala	Val	Ser	Arg	Pro	Gly	Leu	Gly	Lys	Pro	Arg	Phe	Ile
	K ^b	-	-	-	-	-	-	-	-	Val	-	-	-	-	-	-	-	-	-	Glu	-	-	Tyr	Met
	K ^d	-	-	-	-	-	-	-	-	Val	-	-	-	-	-	-	-	-	-	Glu	-	-	-	-
										30	40													
	K ^k	Ser	Val	Gly	Tyr	Val	Asp	Asp	Thr	Gln	Phe	Val	Arg	Phe	Asp	Ser	Asp	Ala	Glu	Asn	Pro	Arg	Tyr	Glu
	K ^b	Glu	-	-	-	-	-	-	-	Glu	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	K ^d	Ala	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Asp	-	-	-	Phe	-
										50	60													
	K ^k	Pro	Arg	Val	Arg	Trp	Met	Glu	Gln	Val	Glu	Pro	Glu	Tyr	Trp	Glu	Arg	Asn	Thr	Gln	Ile	Ala	Lys	Gly
	K ^b	-	-	Ala	-	-	-	-	-	Glu	Gly	-	-	-	-	-	-	Glu	-	-	Lys	-	-	-
	K ^d	-	-	Ala	Pro	-	-	-	-	Glu	Gly	-	-	-	-	-	Glu	Gln	-	-	Arg	-	-	Ser
										70	80											90		
	K ^k	Asn	Glu	Gln	Ile	Phe	Arg	Val	Asn	Leu	Arg	Thr	Ala	Leu	Arg	Tyr	Tyr	Asn	Gln	Ser	Ala	Gly		
	K ^b	-	-	-	Ser	-	-	-	Asp	-	-	-	Leu	-	Gly	-	-	-	-	-	Lys	-		
	K ^d	Asp	-	-	Trp	-	-	-	Ser	-	-	-	Gln	-	-	-	-	-	-	-	Lys	-		
α2-DOMAIN										100	110													
H-2	K ^k	Gly	Ser	His	Thr	Phe	Gln	Arg	Met	Tyr	Gly	Cys	Glu	Val	Gly	Ser	Asp	Trp	Arg	Leu	Leu	Arg	Gly	Tyr
	K ^b	-	-	-	-	Ile	-	Val	Ile	Ser	-	-	-	-	-	-	-	Gly	-	-	-	-	-	-
	K ^d	-	-	-	-	-	-	-	-	Phe	-	-	Asp	-	-	-	-	-	-	-	-	-	-	-
										120	130													
	K ^k	Glu	Gln	Tyr	Ala	Tyr	Asp	Gly	Cys	Asp	Tyr	Ile	Ala	Leu	Asn	Glu	Asp	Leu	Lys	Thr	Trp	Thr	Ala	Ala
	K ^b	Gln	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	K ^d	Gln	-	Phe	-	-	-	-	Arg	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
										140	150													
	K ^k	Asp	Met	Ala	Ala	Leu	Ile	Thr	Lys	His	Lys	Trp	Glu	Gln	Ala	Gly	Asp	Ala	Glu	Arg	Asp	Arg	Ala	Tyr
	K ^b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Glu	-	-	Leu	-	-	
	K ^d	-	Thr	-	-	-	-	-	Arg	Arg	-	-	-	-	-	-	-	-	-	-	Tyr	Tyr	-	-
										160	170											180		
	K ^k	Leu	Glu	Gly	Thr	Cys	Val	Glu	Trp	Leu	Arg	Arg	Tyr	Leu	Gln	Leu	Gly	Asn	Ala	Thr	Leu	Pro	Arg	Thr
	K ^b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Lys	Asn	-	-	-	-	Leu	-	
	K ^d	-	-	-	Glu	-	-	-	-	-	-	-	-	-	-	Glu	-	-	Glu	-	-	Leu	-	
α3-DOMAIN										183	190											200		
H-2	K ^k	Asp	Ser	Pro	Lys	Ala	His	Val	Thr	Arg	His	Ser	Arg	Pro	Glu	Asp	Lys	Val	Thr	Leu	Arg	Cys	Trp	Ala
	K ^b	-	-	-	-	-	-	-	-	His	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	K ^d	-	-	-	-	-	-	-	-	Tyr	-	Pro	-	Ser	Gln	Val	Asp	-	-	-	-	-	-	-
										210	220													
	K ^k	Leu	Gly	Phe	Tyr	Pro	Ala	Asp	Ile	Thr	Leu	Thr	Trp	Glu	Leu	Asn	Gly	Glu	Glu	Leu	Thr	Gln	Asp	Met
	K ^b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Ile	-	-	
	K ^d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Asp	-	-	-	
										230	240											250		
	K ^k	Glu	Leu	Val	Glu	Thr	Arg	Pro	Ala	Gly	Asp	Gly	Thr	Phe	Gln	Lys	Trp	Ala	Ser	Val	Val	Val	Pro	Leu
	K ^b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	K ^d	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Ala	-	-	-	
										260	270													
	K ^k	Gly	Lys	Glu	Gln	Tyr	Tyr	Thr	Cys	His	Val	Tyr	His	Gln	Gly	Leu	Pro	Glu	Pro	Leu	Thr	Leu	Arg	Trp
	K ^b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	K ^d	-	-	-	-	Asn	-	-	-	-	-	His	-	Lys	-	-	-	-	-	-	-	-	-	



The codons used here for tyrosine (TAT) and asparagine (AAT) differ by only one nucleotide, most likely changed by a single mutation.

Identification of K allele-specific sequences

By comparing protein and nucleotide sequences of the three H-2 genes we have identified sequences unique for each K allele. Allele specific sequences are very useful as such information can be used for the synthesis of unique oligomeric DNA probes. We have found one such sequence for the H-2K^k gene in the α 1 domain, amino acids 62-66 (Figure 5A). This sequence is 15 nucleotides long and differs by 7 and 5 nucleotides from the H-2K^d and K^b alleles, respectively. A K^b specific sequence is found in the α 2 domain, residues number 94-100 (Figure 5B). This sequence is 19 nucleotides long and has already been used as probe to identify K^b specific sequences (30). Residues 191-198 of the α 3 domain of the K^d antigen define a sequence potentially unique for the K^d allele. It is distinct from the K^k and K^b sequences in 9 and 8 nucleotides, respectively out of 24 (Figure 5C).

DISCUSSION

The DNA containing the authentic H-2K^k gene was isolated from the two inbred mouse strains B10.BR and AKR/J. The B10.BR DNA was found to contain a second H-2K gene. This clone of DNA (7A) was shown to contain an H-2K^b-like gene (B. Arnold, A. Archibald and S. Kvist, unpublished results). When introduced into 1T 22-6 cells (H-2^q), a fibroblast cell line, the DNA of clone 27-2 (AKR/J DNA) directed the expression of the H-2K^k antigen. This was demonstrated by using monoclonal antibodies specific for the H-2K^k antigen (23). We believe that clone 6D contains the 3' end of the H-2K^k gene of the B10.BR strain for the following reasons. First, it

Figure 4. Amino acid comparison of the three extracellular domains of three H-2K allelic antigens.

On top the sequence of the H-2K^k allele is shown. Homology is indicated by a dash. The boxed amino acids are allele-specific. The two stars indicate the glycosylation sites and the arrow at position 256 denotes the additional glycosylation site of the H-2K^d antigen.

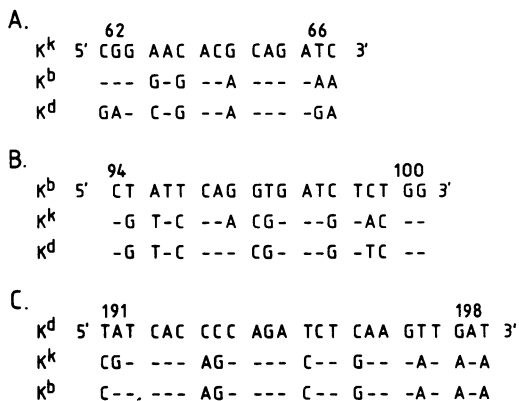


Figure 5. Allele specific sequences for the H-2K locus. Dashes indicate homology to the sequence shown on top. The numbers refer to the amino acid positions in the antigen.

carries the K-locus specific sequence at its 3' end and can therefore be assigned to the K-locus. Secondly, its restriction map is identical to that of clone 27-2, which contains the authentic H-2K^k gene. Thirdly, partial DNA sequence analysis of clone 6D confirms its identity with the gene in clone 27-2.

The DNA sequence of the K^k gene of clone 27-2 was determined and shows an extraordinarily high degree of homology to the H-2K^b gene. The differences are almost exclusively confined to exons 2 and 3. It has been shown that the introns of the K^d and K^b genes display a higher degree of homology than the exons of these genes (11). The same is true when these genes are compared to the H-2K^k sequence. The introns of all three H-2K alleles are 95-98% homologous whereas the corresponding figure for the exons is 92-94%.

Recently, the complete amino acid sequence of the $\alpha 1$ domain of the K^k antigen has been determined (31). This sequence is identical to the sequence deduced from the K^k gene presented here, except for one amino acid. Comparison of the amino acid sequences of the three H-2K alleles reveals that the K^k and K^b antigens are much more similar to each other than to the K^d antigen. The differences between the two former proteins are con-

fined to the $\alpha 1$ and $\alpha 2$ domains where the nucleotide sequences also have diverged most extensively.

In nine amino acid positions of the three extracellular domains we have found unique residues for each K allele (Figure 4). Of these 27 amino acids not less than 14 are charged or polar residues (Arg, Lys, Asn, Asp, Gln, Glu and His) and as many as 21 are clearly nonhydrophobic residues (i.e. when serines and tyrosines are included). We would expect these amino acids to be exposed on the surface of the protein and thereby accessible to interact with other molecules. Future studies, using site directed mutagenesis, will clarify whether these residues are directly involved in antigenic determinants for antibodies or cytolytic T cells or not.

We have identified specific nucleotide sequences for each of the H-2K alleles. Such sequences are useful as synthesized oligonucleotides can be used as probes to distinguish one H-2 gene from another. The K^b allele-specific sequence is unique with respect to both locus and haplotype (30). For the H-2 K^k and K^d allele-specific sequences (Figure 5) we do not know if other H-2 genes within the same haplotype ($H-2^k$ and $H-2^d$, respectively) carry similar or identical sequences. If this would be the case it might be possible to use the allele-specific probes to study the question of whether one of these genes (for instance in the D, Qa or Tla regions) has donated the unique sequence to the K gene. In such a study the donor gene for the K^{bm1} mutant gene has already been identified (32,33,34,35). Further analyses with such probes might help elucidate the mechanisms(s) of the phenomenon termed gene conversion.

ACKNOWLEDGEMENTS

We thank M. Steinmetz for the cosmid 27-2 containing the H-2K^k gene of the AKR/J strain. We are grateful to A.-M. Frischauf and H. Lehrach for the gift of the B10.BR DNA library. For technical assistance we thank G. Küblbeck and C. Raynoschek. For excellent help in preparing the manuscript we thank I. Benner, H. Seifert and C. Droz. H.-G.B. was supported by the grant Do

199 14-2 from the Deutsche Forschungsgemeinschaft. A.L.A. was supported by the Agricultural and Food Research Council, London.

+Present address: German Cancer Research Center, Im Neuenheimer Feld 280, D-6900 Heidelberg, FRG

§Present address: Swiss Institute for Experimental Cancer Research, Ch. des Boveresses, CH-1066 Epalinges s/Lausanne, Switzerland

*To whom correspondence should be addressed

REFERENCES

1. Klein, J. (1975) Biology of the mouse histocompatibility complex. Springer-Verlag, Berlin.
2. Klein, J. (1979) Science 203,516-521
3. Coligan, J.E., Kindt, T.J., Uehara, H., Martinko, J. and Nathenson, S.G. (1981) Nature 291,35-39
4. Zinkernagel, R.M. and Doherty, P.C. (1974) Nature 248,701-702
5. Zinkernagel, R.M. and Doherty, P.C. (1974) Nature 251,547-548
6. Zinkernagel, R.M. and Doherty, P.C. (1979) Adv. Immunol. 27, 51-177
7. Steinmetz, M., Frelinger, J.K., Fisher, D., Hunkapiller, T., Percira, D., Weissman, S.M., Uehara, H., Nathenson, S.G. and Hood, L. (1981) Cell 24, 125-134
8. Evans, G.A., Margulies, D.H., Camerini-Otero, R.D., Ozato, K. and Seidman, J.G. (1982) Proc. Natl. Acad. Sci. USA 79,1944-1998.
9. Moore, K.W., Sher, B.T., Sun, Y.H., Eakle, K.A. and Hood, L. (1982) Science 215,679-682.
10. Kvist, S., Roberts, L. and Dobberstein, B. (1983) EMBO J. 2, 245-254
11. Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrner, K., Kvist, S. and Flavell, R.A. (1983) EMBO J. 2,453-462
12. Steinmetz, M., Winoto, A., Minard, K. and Hood, L. (1982) Cell 28, 489-498.
13. Blobel, G. and Dobberstein, B. (1975) J. Cell Biol. 67,835-851
14. Steinmetz, M. and Hood, L. (1983) Science 222, 727-733
15. Murray, N.E. (1983) In Lambda II, Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A., Eds., pp.395-432, Cold Spring Harbor Laboratory
16. Benton, W.D. and Davis, R.W. (1977) Science (Wash.) 196,180-182
17. Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc. Natl. Acad. Sci. USA 77,5172-5176
18. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387-2399
19. Frischauf, A.-M., Garoff, H. and Lehrach, H. (1980) Nucl. Acids Res. 8,5541-5549
20. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560

21. Wigler, M., Pellicer, A., Silverstein, S., Axel, R., Urlaub, G. and Chasin, L. (1979) *Proc. Natl. Acad. Sci. USA* 76,1373-1376
22. Southern, P.J. and Berg, P. (1982) *J. Mol. App. Gen.* 1,327-341
23. Arnold, B., Burgert, H.-G., Hamann, U., Hämmerling, G., Kees, U. and Kvist, S. (1984) *Cell* 38,79-87
24. Kvist, S., Wiman, K., Claesson, L., Peterson, P.A. and Dobberstein, B. (1982) *Cell* 29,61-69
25. Lemke, H., Hämmerling, G.J. and Hämmerling, U. (1979) *Immunol. Rev.* 47,175-206
26. Lallanne, J.L., Bregegere, F., Abastado, J.P., Gachelin, G. and Kourilsky, P. (1982) *Nucleic Acids Res.* 10,1039-1049
27. Xin, J.-H., Kvist, S. and Dobberstein, B. (1982) *EMBO J.* 1, 467-471
28. Steinmetz, M., Malissen, M., Hood, L., Drn, A., Maki, R.A., Oastoornikoo, G.R., Stepan, D., Gibb, E. and Romaniuk, R. (1984) *EMBO J.* (in press)
29. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383
30. Schulze, D.H., Pease, L.R., Obata, Y., Nathenson, S.K., Reyes, A.A., Ikuta, S. and Wallace, R.B. (1983) *Mol. and Cell Biol.* 3,750-755
31. Lillehoj, E.P. and Coligan, J.E. (1984) *Mol. Immunol.* 21,185-190
32. Pease, L.R., Schulze, D.H., Pfaffenbach, G.M. and Nathenson, S.G. (1983) *Proc. Natl. Acad. Sci. USA* 80,242-246
33. Schulze, D.H., Pease, L.R., Geier, S.S., Reyes, A.A., Sarmiento, L.A., Wallace, R.B. and Nathenson, S.G. (1983) *Proc. Natl. Acad. Sci. USA* 80,2007-2011
34. Weiss, E.H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J. and Flavell, R.A. (1983) *Nature* 301,671-674
35. Mellor, A.L., Weiss, E.H., Ramachandran, K. and Flavell, R.A. (1983) *Nature* 306,792-795.